# Supporting Information

## Jordan et al. 10.1073/pnas.1308282110

### SI Experimental Procedures

**Media and Strains.** Strains were maintained in duplicate at 16 °C in 10 mL soybean cultures (1) and subcultured every 6 mo. Frozen stocks were concurrently maintained. Forty-eight hours before an experiment, 50 μL of a soybean culture was used to inoculate 50 mL of the appropriate experimental growth medium. The base growth medium is supplemented proteose peptone (SPP) medium (1xR, "1× Rich"): 2% proteose peptone, 0.1% yeast extract, 0.2% glucose, and 0.003% of the iron compound EDTA ferric sodium salt (wt/vol) (Sigma E6760). Cultures were grown at room temperature without shaking for 48 h before each experiment in the appropriate medium. Media used in this study are abbreviated in the main text as follows:1× SPP, 1xR; 2× SPP, 2xR; 1xR sterile filtered plus polymer beads, 1xB; Bacterized, Bac; chemically defined medium, CDM.

The growth medium 2xR is 1xR with twice the concentration of each ingredient (4% proteose peptone, 0.2% yeast extract, 0.4% glucose, and 0.006% iron) (2). The growth medium 1xB is the 1xR medium that has been filtered with a 0.2-μm sterile filter and then supplemented with 1.57-μm diameter polymethyl methacrylate beads (Bangs Laboratories; P0015700PN) at a density of 6,800 beads/mm$^3$ or $\sim 3.2\times10^4$ beads per chamber (3). Bacterized medium was prepared by growing *Escherichia coli* (DH5α) in 1xR media to an optical density of 0.6 before sterilizing via an autoclave. Chemically defined medium was taken directly from Szablewski et al. (4).

The two strains used in this study were *Tetrahymena borealis* (SD01609) and *Tetrahymena thermophila* CU428.2 SD00178. All strains were obtained from the Cornell *Tetrahymena* Stock Center, http://tetrahymena.vet.cornell.edu.

**Microscopy.** ***Imaging.*** The imaging apparatus consists of a light-emitting diode light source (LuxDrive; 5027 white), a condenser lens ($f$ = 35 mm), a sample stage, a focusing relay lens (NT45-762; Edmund Optics), and an imaging sensor. The image sensor is 1,600 × 1,200 pixels and was taken from a consumer webcam (Logitech Quickcam 9000) that acquires images at 15 Hz. The pixel size of this sensor is 33 μm. An image of the sample plane is focused onto the sensor, using the 1× relay lens. In the arrangement used here the magnification of the optical system is 0.78, giving a pixel size in the object plane of $4.25\pm0.04$μm determined using a US Air Force test target (Edmund Scientific). Movies were recorded using the commercially available Logitech software, compressed, and stored to a disk to be processed at a later time.

***Imaging uncertainty.*** Uncertainty in the position as a result of the imaging hardware (e.g., illumination variation, pixel noise, or optical point-spread function) was evaluated using a USAF test target. The movie was recorded for ∼1 min and the resulting 1,000 frames were processed using the same custom MATLAB (Mathworks) algorithms developed for tracking (*Tracking* section). The centroid positions of the objects from the resulting segmented images were recorded and their deviations across the 1,000 frames were measured to be on average 0.2 μm, with a maximum of 1 μm. There was no clear dependence on this variance with the size of the objects between 10 and 4,500 square pixels.

***Temperature control.*** Each microscope was assembled on an optical rail and mounted to an optical breadboard. Four microscopes were housed in a single light tight, thermally insulated box. The air in the box was circulated with a fan. The temperature was controlled using a custom feedback system, using a Peltier heating/cooling element. The thermometer, located near the sample stages, is a linearized thermistor (Omega Engineering; 44204) with a custom amplifier (5). The thermometer amplifier output is digitized via a LabJack U3 USB DAQ (Labjack; U3-HV) interface with a precision of 1.2 mV (4.8 mK). This temperature is controlled with a proportional integral (PI) feedback loop implemented in MATLAB. This PI feedback stabilizes the measured temperature to the set point with SD of 4.6 mK from the set point across all experiments. The thermometer is calibrated to ±50 mK absolute accuracy. All of the experiments presented here were carried out at 23 °C.

***Chamber fabrication.*** Microfluidic chambers like the one shown in Fig. 1*A* of the main text were constructed from polydimethylsiloxane (PDMS), using conventional soft lithography techniques (6). SU8-2075 (Microchem) negative photoresist was spin-cast onto a 4-inch diameter silicon wafer. Four milliliters of photoresist was dispensed onto the center of the wafer and the wafer was spun at 500 rpm for 10 s and then at 1,000 rpm for 30 s to a thickness of 240 μm (Laurell, WS-600-6NPP spin coater). The soft bake was 7 min at 65 °C followed by 45 min at 95 °C. The wafer was then masked with a printed transparency mask (Pageworks) and exposed to 575 mJ of radiation over 18 s and baked postexposure for 5 min at 65 °C and 15 min at 95 °C. The resulting mold was developed for 15 min in SU-8 developer. This mold was exposed to tridecafluoro-1,1,2,2 tetrahydrooctyl trichlorosilane for 1 h in mild vacuum to prevent sticking of PDMS. Sylguard 184 Silicone Elastomer (Dow Corning) was mixed with Sylguard 184 curing agent at a ratio of 8:1 and poured onto the silicone mold. This was cured for 25 min at 80 °C and then allowed to cool. The PDMS chamber was trimmed using a razor blade and holes were punched using a syringe needle. The chamber was then oxygen plasma treated (Harrick plasma cleaner) and sealed to a glass coverslip before being cured for an additional 1 h at 80 °C. The chambers used here were 4.9 mm in diameter and 240 μm deep for a total volume of 4.5 μL.

The single-layer chamber design used in this work is shown in Fig. 1*A*. The central ring in the chamber is masked during plasm bonding and is therefore not adhered to the coverslip. As a result, when pressure is applied to the input, the chamber distends upward, allowing medium and organism to flow beneath. When a single cell is observed in the chamber, the pressure is released, and the ring of PDMS settles on the coverslip and traps the cell in the chamber. Movie recording is then initiated.

***Nutrient exhaustion.*** The volume of a single *T. thermophila* is approximately 1 millionth the volume of media in the chamber. We have estimated the rate of nutrient uptake based on the work of Andersen et al. (2) to be about $10^{-12}$g/min. Over the course of the experiment, we therefore estimate on the order of $10^{-9}$g to be consumed. The chamber starts with $10^{-4}$g of material; this indicates that only 0.001% of the nutrients are used over the course of the experiment.

**Tracking.** ***Image segmentation.*** To segment swimming cells from stationary background we used a dynamically constructed background image. Each image is captured and imported into MATLAB, using the VideoIO toolbox (Gerald Daley, http://sourceforge.net /projects/videoio/) as a matrix $I_{ij}(t)$ with $1\le i\le1{,}200$ and $1\le j\le1{,}600$, where $I_{ij}(t)$ represents the value of the intensity of the pixel at location $(i,j)$ at time $t$. A dynamic background image ($B$) is computed with $B_{ij}=max(I_{ij}(t),I_{ij}(t+\tau))$, where $\tau$ is set empirically to be 200 s. Thus, objects that do not move for 200 s will not be detected. We find that *Tetrahymena* generally does

not remain stationary for longer than this period. The background subtracted image is segmented for moving objects, using a global threshold that was empirically determined to be 0.2 of the maximum pixel value in an image. In this manner, artifacts due to long-timescale changes in the image over the course of the experiment can be avoided, such as changes in illumination or shifts of the imaged volume.

Moving objects are filtered to retain only the expected number $N_0$ in the expected size range of the *Tetrahymena* (>25 square pixels). For each retained segmented object the centroid, area, orientation, and eccentricity are recorded using the regionprops function in the MATLAB Image Processing Toolbox.

***Tracking.*** We constructed swimming trajectories from segmented images with custom-written MATLAB code that follows closely a method developed by Jaqaman et al.(7). Trajectories are created by matching objects in frame $t$ with those in $t+1$. This problem can be posed as a linear assignment problem. Each object $x_i$ in frame $t$ can be linked to an object $x_j$ in frame $t+1$ for a cost $c_{link}(x_i,x_j)$. The cost function used is the sum of the Euclidian distance between objects and a weighted difference in area. Alternatively each $x_i$ can remain unlinked in frame $t+1$ for a cost $c_{lose}$ and $x_j$ can remain unlinked to an object in frame $t$ for a cost $c_{find}$. The total cost of an assignment for all blobs between two frames is given by $C=\sum_{ij}\hat{A}_{ij}c_{ij}$, where $\hat{A}_{ij}$ is the assignment matrix with $\hat{A}_{ij}=1$ if $x_i$ is linked to $x_j$ and 0 otherwise. We use the Jonker–Volgenant algorithm (8) to solve the linear assignment problem and find the assignment matrix $\hat{A}_{ij}$ that has the minimal total cost of assignment between two frames.

***Cost matrix construction:*** The list of $N$ detected particles in frame $t$ is given as $X^N(t)=\{x_i(t)\}$ for $i=1:N$. For tracking we seek to match these $N$ particles to the $M$ particles detected in frame $t+1$ given by $X^M(t+1)$. Here we define the costs for linking any two particles between frames ($c_{link}$) and for leaving particles unlinked ($c_{lose}$, $c_{find}$) and provide the construction of the cost matrix. Because particles may disappear and artifacts may arise during object detection, $N$ does not always equal $M$. We have used the cost matrix formulation of Jaqaman et al. (7) to account for these possibilities. The cost matrix $C$ is constructed from the actual $N\times M$ linking matrix $c_{ij}=c_{link}(x_i(t),x_j(t+1))$ in the following way,

$$C=\begin{vmatrix} c_{ij} & c_{lose}(X^N(t))I_N \\ c_{find}(X^M(t+1))I_M & c_{ij}^T \end{vmatrix},$$

which is an $(N+M)\times(N+M)$ matrix, where $c_{lose}$ is the cost of an object in frame $t$ remaining unassigned, $c_{find}$ is the cost of an object in frame $t+1$ remaining unassigned, and $I_n$ is the $n\times n$ identity matrix. The $c_{ij}^T$ matrix is added to satisfy the requirements of the linear assignment formulation. We empirically determined the values for finding and losing a particle in frame $t$ as

$$c_{lose}=c_{find}=1.25\times max\left(\left\{\hat{A}_{ij}(t')*C(t')|t'=1:t,i<N,j<M\right\}\right),$$

where $N=|X^N(t)|$ and $M=|X^M(t+1)|$, and $i<N$ and $j<M$ ensure only linking assignments are considered.

***Linking costs:*** The frame-to-frame linking cost function is given by

$$c_{link}(x_i(t),x_j(t+1))=\Delta r(x_i(t),x_j(t+1))+\alpha\Delta A(x_i(t),x_j(t+1)),$$

where $\Delta r$ is the displacement of the centroids of the objects and $\Delta A$ is the difference in area and $\alpha$ is a scaling factor. In our case, $\alpha$ is set to 0.2. For the segment linking step, let $S^N$ be the set of $N$ segments $s_i$ to be joined. The cost function to join segment $s_i$ to $s_j$, $i\neq j$, is

$$c_{link}(s_i,s_j)=\Delta r_{ij}(s_i,s_j)+\tau\Delta t_{ij}(s_i,s_j)-\Delta v_{ij},$$

where $\Delta r_{ij}$ is the displacement from the end of segment $i$ to the beginning of segment $j$, $\Delta t_{ij}$ is the time of the gap in frames, $\tau$ is a scaling factor (0.5 in this case), and $\Delta v_{ij}$ is the projection of the velocity in the last frame of $s_i$ onto the velocity in the first frame of $s_j$, normalized to $\Delta r_{ij}$.

***Tracking fidelity.*** Correct assignments are essential to maintain the identity of individuals over the course of the experiment. In general, when the ratio of the frame-to-frame displacement and the interobject distance is ~1 and each is small (<212 μm), there is the possibility of an erroneous assignment. Sets of contiguous frames in which this ratio is close to 1 are called "crossover events". This ratio is a function of the average speed of the objects, the frame rate of the movie, and the density of objects. In an average experiment, there are ~25 crossover events per object pair per "lifetime" in which the above ratio is <1.6. Across all experiments, 96.2% of the crossover events are correctly assigned by the tracking algorithm. All crossover events are inspected manually and those that are not correctly assigned are corrected by hand. The validity of the hand-scored trajectory is verified by an analysis of the movie and comparison of the resulting lifetimes of individuals, the speed of each, the angular acceleration, and, importantly, the cell size before and after the crossing. As a result, we can easily maintain identity with up to eight individuals in the chamber.

**Data Analysis.** ***Speed and angular velocity.*** The trajectories resulting from our tracking algorithm consist of $(x,y)$ coordinates for each cell as a function of time at a frequency of 15 Hz. As discussed in the main text, we transform trajectories into a time series of speed and turning angle. Let $\mathbf{v}=(v_x,v_y)=((\Delta x/\Delta t),(\Delta y/\Delta t))$. Speed is then $\mathbf{v}|$ calculated between pairs of frames separated by $\Delta t=1/15$s, and angular velocity is calculated between pairs of velocity vectors by

$$\omega(t)=\frac{sign(\mathbf{v}(t+1)\times\mathbf{v}(t))}{\Delta t}cos^{-1}\frac{((\mathbf{v}(t+1)\cdot\mathbf{v}(t))}{(|\mathbf{v}(t+1)||\mathbf{v}(t)|)}. \quad \textbf{[S1]}$$

From $|\mathbf{v}(t)|$ and $\omega(t)$ we construct 2D histograms ($P(\omega,|\mathbf{v}|)$) as discussed below.

***Boundary effects.*** The confined two-dimensional geometry of the chambers used for these experiments requires that we investigate the effects of interactions with the boundaries.

First, we study the effects of a cell's proximity to the radial boundary, or the "wall", of the chamber (Fig. S1*A*, *Inset*). Fig. S1*B* indicates that beyond a distance of ~ 40 μm from the wall of the chamber the density of segmented blobs is uniform. This naturally defines a wall interaction region given by an annulus that is 40 μm wide with its outer edge at the wall. We find that the fraction of the total lifetime that an individual spends within this region varies between 5% and 30% across conditions as shown in Fig. S1*C*. Fig. S1*B* shows that there is no lasting, measurable, effect on the speed and turning angle distributions of these wall interaction events.

However, $P(\omega,|\mathbf{v}|)$ distributions [$P(|\mathbf{v}|,\omega)$ *histograms* section] differ significantly between cells swimming within the wall interaction region and those swimming the bulk of the chamber. The effect of these differences on the quantities calculated here is shown in Fig. S2 *B* and *C*.

Next, we study the effect of interactions between the "ceiling" and the "floor" of the chamber. These surfaces are perpendicular to the optical axis of the microscope and composed of PDMS and glass, respectively. Fluid dynamic studies of *Tetrahymena* in confined spaces have been previously conducted by H. Winet (9). These results indicate that the velocity of a swimming cell should not be altered by confinement to a liquid layer of thickness four

times the radius of the cell or ∼80 μm [figure 6 and table 4 in Winet (9)]. We have confirmed this prediction by making measurements of swimming *T. thermophila* in chambers of various depths. Three sets of experiments (18 individuals per experiment) were done at three chambers depths, 45 μm, 85 μm, and 240 μm. We see an effect in peak swimming speed, which is the same for 85- and 230-μm geometries, but decreased for 45-μm geometry (Fig. S1*E*). We conclude that for the swimming we observe in the 240-μm–deep chambers, where our data were taken, the cells' swimming behavior is not significantly perturbed by their interactions with the floor or the ceiling of the chamber.

***P*(|v|, ω) histograms.** The construction of $P(|\mathbf{v}|,\omega)$ histograms is done in one of two ways. To measure changeability and memory, where absolute time is considered, a trajectory is divided into nonoverlapping windows of fixed width ($t_w$). We denote these histograms $P^N(t)$, where $N$ is the index of the individual and $t$ the time since the beginning of the lifetime in minutes (Fig. 1*D* and Fig. 2*B*). We set the nonoverlapping window size to 2,000 frames to maximize the temporal resolution while keeping finite sampling effects to a minimum. With this window size the bias and variance in our estimates of the Jensen–Shannon divergence are small (∼0.01, *Bias* and *Variance* sections) and the timescale of 2.2 min is short compared with the lifetime of an individual of 2–6 h, but fast compared with the behavioral dynamics we observe (Fig. 2 *A*–*C*, main text).

For measurements of individuality (and therefore plasticity and heritability) we divide trajectories into a constant number of bins (100 in all data presented in the main text). In these calculations $t_w$ varies from one individual to the next, and a distribution of $t_w$ values by condition for wild-type (WT) *T. thermophila* individuals is shown in Fig. S2*A*. For these histograms time is normalized to 1 and indexed by $s$ ($P^N(s)$). This was done because the computation of individuality matrices is costly, scaling as $O(n^2)$, where $n$ is the number of segments, and this becomes prohibitive for $n > 2\times10^4$. Multidimensional scaling (MDS) embeddings are computationally expensive as well, scaling as $O(n^2 d_e)$, where $d_e$ is the dimensionality of the embedding space. Because embeddings are calculated for large individuality matrices (171 individuals), and repeatedly, to measure the sensitivity of the process to errors, it is prohibitive to do so for individuality matrices with hundreds of $P^N(s)$ histograms per individual. We therefore chose 100 as the largest number of segments per individual that was computationally tractable while keeping the maximum $t_w$ used shorter than the behavioral memory.

In subdividing the trajectory into nonoverlapping windows we lose information regarding the dynamics on timescales shorter than $t_w$ minutes and assume that behavior is not dependent on the location of the individual in the PDMS chamber. Our data indicate that the latter assumption is true except at the boundaries of the chamber. To show this we examined the spatial correlation between individuals in the same chamber and compared it to the spatial correlation of individuals in two different chambers. If there were long-lasting spatial heterogeneities that resulted in some "preferred" locations within the chamber, this would be evident as a greater spatial correlation between individuals in the same chamber. We observed no such difference (Fig. S1*D*), indicating no large-scale or long-lasting spatial heterogeneity. We also find that excluding the behavior from points in time where the individual is in close proximity to the wall does not alter our estimates of changeability or individuality significantly. We find that the interactions with the boundaries do not significantly alter our measurements of behavioral similarity (Fig. S2).

***Jensen–Shannon divergence.*** As discussed in the main text, we compare $P(|\mathbf{v}|,\omega)$ distributions [$P^N(t)$ and $P^N(s)$], using the Jensen–Shannon divergence, which for two distributions $p$ and $q$ is given by

$$D(p|q) = \frac{1}{2}\left(\sum_i\left(p_i log\frac{p_i}{m_i}\right) + \sum_i\left(q_i log\frac{q_i}{m_i}\right)\right). \qquad \textbf{[S2]}$$

here $m_i = (p_i + q_i)/2$ and $i$ is an index over all bins in the histogram. To construct histograms we use the Freedman–Diaconis method to determine the bin size (10). Histogram binning is specified dynamically with this method for each pair of histograms in a changeability or individuality matrix.

***Statistical properties of divergence estimates.*** Given two histograms $\mathbf{h}^1$, $\mathbf{h}^2$, which are estimates of distributions $\mathbf{p}^1$, $\mathbf{p}^2$, we want to quantify the bias and variance of our estimates of $D(\mathbf{p}^1|\mathbf{p}^2)$ given by $D(\mathbf{h}^1|\mathbf{h}^2)$. Analytic expressions for the bias and uncertainty are given in Grosse et al. (11).

***Bias.*** The bias is given as the difference between the expected value of $D(\mathbf{h}^1|\mathbf{h}^2)$ and that of $D(\mathbf{p}^1|\mathbf{p}^2)$. Following Grosse et al. (11), let us consider two sequences of observations of length $N$, each of which can be in $i = 1...B$ "states" according to the unobserved distributions $\mathbf{p}^1$ and $\mathbf{p}^2$. In our work, these states are bins in our histograms. Let $n_i$ represent the number of observations in each state $i$. Thus, our histograms $\mathbf{h}^{1,2}$ are given by $\left\{\frac{n_i}{N}\right\} i = 1...B$:

$$bias = \left\langle D\left(\mathbf{h}^1|\mathbf{h}^2\right)\right\rangle - D\left(\mathbf{p}^1|\mathbf{p}^2\right).$$

The bias is independent of the underlying distributions $\mathbf{p}^\alpha$ ($\alpha = \{1,2\}$) and depends only on the ratio of the number of occupied states $B$ and the number of observations $N$. State $i$ is occupied if $\mathbf{p}_i \neq 0$. The bias is then given by

$$\left\langle D\left(\mathbf{h}^1|\mathbf{h}^2\right)\right\rangle - D\left(\mathbf{p}^1|\mathbf{p}^2\right) = \frac{B^* - 1}{2Nln(2)}.$$

The bias represents the average divergence one would expect to measure between independent samples drawn from identical distributions. This value is systematically biased upward from its true value due to sampling error and decreases with better sampling (larger $N$). In this work, $B^* \sim 40$ and $N \geq 1{,}000$, giving a bias of 0.02 bit. We have confirmed this bias estimate by performing numerical simulations on uniform distributions, for which $D_{JS}$ can be calculated analytically. In this regime we find that the bias correction performs well even when $N \sim 3B$. Combing the analytical bias correction with the optimal binning strategy further reduces the bias to <3% even for modest values of $N$. Therefore, we use both this analytical bias estimate and an optimal binning strategy for correcting our measurement of $D_{JS}$ on each pair of histograms.

***Variance.*** The uncertainty is given as the variance of $D(\mathbf{h}^1|\mathbf{h}^2)$:

$$uncertainty = V\left[D\left(\mathbf{h}^1|\mathbf{h}^2\right)\right]. \qquad \textbf{[S3]}$$

Grosse et al. (11) show that the uncertainty $V[D(\mathbf{h}^1|\mathbf{h}^2)]$ depends on the number of observations, $N$, and depends only on terms of $O(1/N^2)$ and smaller. However, a quantitative estimate of the variance in $D_{JS}$ is important for measuring the significance of the stress in our MDS embeddings (*Multidimensional scaling analysis* section). We have accomplished this by bootstrapping. To do this we randomly chose 5,000 pairs of histograms from the large individuality matrix used to construct the MDS embeddings in Fig. 3 in the main text (171 individuals, 100 histograms per individual). For each of these pairs of histograms we compute 500 bootstrapped histograms $D_{JS}$ between each pair. We find $\sigma^{100}_{D_{JS}} = 0.01$, and this estimate is not strongly dependent on the magnitude of $D_{JS}$. This variance decreases as we increase the number of bins per lifetime (longer $t_w$). Bootstrapping yielded $\sigma^{50}_{D_{JS}} = 0.007$ and $\sigma^{25}_{D_{JS}} = 0.005$.

***Removing divisions.*** For all organisms observed, immediately before division and for a short time after division, swimming ceases or

nearly ceases (Fig. 1*D*). The duration of this period varied from individual to individual and from condition to condition. These periods were included for all analysis except for the analysis of behavioral memory (Fig. 4 *E* and *F*), for which they were removed in the following way. The speed as a function of time was filtered with a median sliding window of 3.3 s. A threshold was then applied at 0.6 of the maximum median filtered speed. All frames before the first crossing and after the last crossing were discarded.

***Multidimensional scaling analysis.*** Multidimensional scaling is a dimensionality reduction technique that looks for a low-dimensional representation of data that preserves distances between points. MDS is well suited to our data because our statistical approach measures the dissimilarity between histograms and not their absolute location in a high-dimensional space. The result is a dissimilarity matrix between behaviors. We apply MDS to this matrix.

For all of the embeddings shown we consider all sublifetime representations of behavior. To do this, we divide each lifetime into $n$ nonoverlapping windows of variable length $t_w$ depending on the lifetime of the individual, $t_w = T_{lifetime}/n$. The resulting individuality matrix is an $M \times M$ dissimilarity matrix with $M = n \times N$, where $N$ is the number of individuals. We have performed MDS embeddings for $n = 25, 50, 100$ and found that our qualitative conclusions are unchanged over this range of $n$. We perform MDS on a dissimilarity matrix $\Delta$ that is symmetric, has zeros on the main diagonal, and entries $\delta_{i,j}$ with $0 \le \delta_{i,j} \le 1$.

We use the MATLAB built-in function "mdscale.m" to implement metric MDS. In metric MDS the user specifies a number of embedding dimensions $d_e$ for the embedding ($d_e$ is less than the rank of the dissimilarity matrix). The goal of the metric MDS optimization is to find the configuration of $M$ points in $d_e$ dimensions whose Euclidean distances correspond as closely as possible to the $\delta_{i,j}$. Briefly, a configuration of $M$ points in $d_e$ dimensions we denote by $X = \{x_1 \ldots x_M\}$, where each $x$ is a vector with size $1 \times d_e$. For a configuration we compute the distance matrix $D$, where entry $d_{i,j} = \sqrt{(x_i - x_j)^2}$. The configuration $X$ is then optimized under the objective function:

$$stress = \sqrt{\frac{\sum_{i<j}\left(\delta_{i,j} - d_{i,j}\right)^2}{\sum_{i<j}\delta_{i,j}^2}}. \qquad \textbf{[S4]}$$

This optimization is accomplished in the "mdscale" implementation by gradient descent. For metric MDS embeddings we used the "metricstress" criterion. For nonmetric MDS embeddings we used the "stress" criterion.

We note that $D_{JS}$ is not a metric because it does not satisfy the triangle inequality. However, metric MDS seeks an embedding of the $D_{JS}$ distances we calculate into a metric Euclidean space. As a result, when we apply metric MDS to $D_{JS}$, we find some residual stress even when embedding in more than six dimensions (Fig. S4*A* and Fig. 3). However, $\sqrt{D_{JS}}$ is a metric (12), so we also performed metric MDS on $\sqrt{D_{JS}}$ for $n = 25, 50, 100$, where we find that the stress continuously decreases toward zero, rather than saturating with a fixed amount of residual stress. This indicates that the residual stress observed in the embedding shown in Fig. 3 results from embedding nonmetric distances ($D_{JS}$) in a Euclidean metric space.

***Number of embedding dimensions.*** There is no general way to determine the intrinsic dimensionality of our data, but in some cases we can estimate the number of dimensions required to accurately represent it. To do this we sought an estimate of the minimum significant change in stress obtained by increasing $d_e$ by 1 ($|\Delta S_{min}|$). In essence, we must compute an uncertainty in the stress and then ask whether increasing the dimensionality of the embedding reduces the stress by more than this uncertainty. If it does not, than adding a dimension is not significantly (in a statistical sense) reducing the stress. In Eq. **S4** $\delta_{i,j}$ has a finite variance due to uncertainty in $D_{JS}$. We propagate this uncertainty to an uncertainty in the stress:

$$|\Delta S_{min}| = \frac{\partial s}{\partial \delta_{i,j}}\sigma_{D_{JS}} = \frac{1}{s}\left(\frac{\sum_{i<j}\left(\delta_{i,j} - d_{i,j}\right)}{\sum_{i<j}\delta_{i,j}^2} - \frac{\sum_{i<j}\left(\delta_{i,j} - d_{i,j}\right)^2 \sum_{i<j}\delta_{i,j}}{\left(\sum_{i<j}\delta_{i,j}^2\right)^2}\right)\sigma_{D_{JS}}. \qquad \textbf{[S5]}$$

In propagating this error we have assumed that $\frac{\partial d}{\partial \delta} \approx 0$. This amounts to assuming that the embedded distances ($d_{i,j}$) do not vary significantly for small changes in $\delta_{i,j}$. For $n = 25$, 50 and $D_{JS}$ we have found numerically that this assumption is a good one, whereas for $\sqrt{D_{JS}}$ this assumption breaks down—that is, the $d_{i,j}$ are changed significantly for small ($\sim 0.01$) changes in the $\delta_{i,j}$. In the latter case errors do not propagate in any simple way. As a result, Eq. **S5** is approximately correct for embeddings performed on the $D_{JS}$, but does not provide a reliable measure of uncertainty in the stress for embeddings done with $\sqrt{D_{JS}}$.

To verify this assumption we performed a bootstrapped estimate of the change in stress arising from $\sigma_{D_{JS}}$ for embeddings with $n = 50$, by adding independent and identically distributed (i.i.d.) Gaussian random numbers with $\sigma = 0.007$ (which corresponds to our bootstrapped estimate of the uncertainty in $D_{JS}$) to the $\delta_{i,j}$ and recomputing the embedding on these perturbed dissimilarity matrices. We found that the analytical expression for the uncertainty in the stress expressed in Eq. **S5** ($|\Delta S|_{min} = 0.002$) compares well to bootstrapped estimates $|\Delta S| = 0.0016 \pm 0.0007$. Bootstrapping for $n = 100$ was found to be too computationally expensive so these dissimilarity matrices were not included in this analysis.

Fig. S4 *C* and *D* shows the reduction in stress from sequentially adding dimensions to the embedding for $n = 25$, 50. For all of these embeddings we find that two to four dimensions are significant as measured by the uncertainty calculated in Eq. **S5**.

***Nontriviality of low-dimensional embeddings.*** We contend that the fact that our behavioral data can be represented in two to three dimensions is nontrivial. Here, we address this in two ways, simulations of synthetic trajectories and analytical limits on dimensionality reduction for random points in high dimensions. First, we used simulations of trajectories constructed from synthetic $P(|v|, \omega)$ histograms (Fig. S5). We constructed three types of mixture distributions with properties similar to those we observe in *T. thermophila*: One simulation generated trajectories from distributions determined by 11 free parameters (3 $\langle|v|\rangle$, 3 $\sigma^2_{|v|}$, 3 $\sigma^2_{\omega}$, and two mixing proportions) (Fig. S5 *A* and *B*), and a second simulation generated the same histograms subject to the hyperbolic constraint $|v| \times \omega < 1$. The hyperbolic constraint mimics a physical constraint by lowering the accessible angular velocities ($\omega$) for rapidly moving cells. This constrained the histograms to have a roaming–dwelling-like structure (Fig. S5 *C* and *D*). Finally, we simulated trajectories from distributions determined by only 3 free parameters ($\langle|v|\rangle$, $\sigma^2_{|v|}$, $\sigma^2_{\omega}$) (Fig. S5 *E* and *F*). For each simulation, trajectories were generated by drawing 2,000 i.i.d. samples from a distribution with randomly chosen means and variances (log-normally distributed) and mixing fractions (uniformly distributed), and this process was repeated 10 times for each simulation. Trajectories from these simulations are shown in Fig. S5 *A*, *C*, and *E*. We then applied our procedure to these trajectories, constructing matrices of $D_{JS}$ and performing metric MDS on those matrices. Fig. S5 *B*, *D*, and *F* shows the stress plots for these embeddings. We find that trajectories constructed from distributions with many free parameters require much higher-dimensional embeddings to be accurately represented. We also find that constraints on the histograms reduce the number of required dimensions for embedding (Fig. S5 *F*

and $G$). This provides qualitative support for our claim that low-dimensional embeddings accurately reflect the dimensionality of $P(|v|,\omega)$ distributions. It also shows that low-dimensional embedding is not a trivial consequence of our data processing scheme or a trivial property of 2D random trajectories.

Second, we consider the following question: Given a random set of 8,550 points with the same dimensionality as our data, how few dimensions would be required to embed those points in a Euclidean space? This question has been considered analytically and bounds on the number of dimensions required to embed random points are known. Here we compare our results for embeddings of $\sqrt{D_{JS}}$ with these bounds. We find that the 2D representation obtained here is well below the number of dimensions required to represent random points, supporting our claim that our embeddings are nontrivial.

Formally, given a collection of $M$ points $\{x_i \ldots x_M\}$, where the distance between them is given by an arbitrary metric $D(x_i,x_j)$ (in our case this would be the JS divergence) gives a metric space $(X,D)$. Let us consider maps from arbitrary metric spaces $(X,D)$ into Euclidean metric spaces, that is, those for which $D$ is the $l_2$ norm, $l_2^d$, where $d$ indicates the dimensionality of the target metric space. Thus, we are looking for a map, such that for every $x_i \in X$, there is a corresponding $f(x_i) \in X'$ such that $D(x_i,x_j) = l_2^d(f(x_i),f(x_j))$.

Such a map is called an embedding. Exact embeddings are not generally possible; however, approximate embeddings can often be found if we allow the distances to be distorted. The expansion of a mapping $f: X \to X'$ is given by

$$\sup_{x_i,x_j \in X} \frac{l_2^d(f(x_i),f(x_j))}{D(x_i,x_j)}$$

and the contraction is the inverse,

$$\sup_{x_i,x_j \in X} \frac{D(x_i,x_j)}{l_2^d(f(x_i),f(x_j))}.$$

The distortion of a mapping is defined as the product of these (13). This brings us to a general result of Bourgain, that any $n$-point metric space $(X,d)$ can be embedded in a Euclidean metric space with distortion bounded by $\log(n)$. However, this bound does not allow us to fix the dimension $d$ of the target space. To do that, we must consider distortions given by $O(n)$ for $d = 2$ and $O(n^{(2/d)} * \log^{(3/2)}(n))$ for $d \geq 3$; see Matousek et al. (14).

In our global embedding, we are representing a set of 8,550 points ($n = 8{,}550$) and the distortion is ~2,000; in comparison, the distortion of a random configuration of points is given by $O(n) \approx 10{,}000$ (this bound is tight), in two dimensions. Thus, this distortion incurred by embedding our data in two dimensions is fivefold lower than expected for a random set of points. For 3D embeddings, we would expect the distortion of a random collection of points to be 12,000, although we find the distortion to be ~970. Clearly our data are not scattered randomly throughout a high-dimensional metric space.

***MDS results and consistency.*** In Fig. 3 of the main text we claim that the two MDS dimensions correspond to changeability and individuality, respectively. Fig. S6 quantifies this relationship. The correlation between the MDS dimensions and the definitions of roaming and dwelling is discussed below (*Definitions of roaming and dwelling* section).

To check that our dimensionality reduction results are not idiosyncratic we performed multiple robustness checks on our conclusions from Fig. 3. We performed both metric and nonmetric MDS on $D_{JS}$ dissimilarity matrices. Nonmetric MDS preserves the rank order of distances without attempting to minimize the differences between $\delta$ and $d$. We performed metric MDS on $\sqrt{D_{JS}}$; in these embeddings formal bounds on dimensionality reduction in metric spaces are applicable (13) (*Nontriviality of low-dimensional embeddings* section). For all of the considered embeddings we found the stress in two dimensions to be low (<0.22). In addition, our interpretation of the first two MDS dimensions, corresponding to changeability and individuality, respectively, was consistent for all of the embeddings we performed (Table S1).

The optimization in of Eq. **S4** is nonconvex. Therefore, we performed repeated embeddings from random initial configurations (for $n = 50$). We found that the stress varied by <1% between embeddings. We observed that 0.1% of the points in the embedding changed their relative position significantly from one embedding to the next. These points corresponded to points with high average stress. Given that the embedding does not change significantly between runs we concluded that our results were not strongly variable between local minima.

Finally, to check that our conclusions were not dependent on comparing $P^N(s)$ using $D_{JS}$ we calculated individuality matrices ($I^{N,M}$), using two other metrics on distributions—the earth mover's distance (EMD) and the Kolmogorov–Smirnov (KS) test statistic on $|v|$ distributions alone. Embedding for both of these distance metrics is shown in Fig. S6 $C$ and $D$ and our interpretations of the dimensions in these embeddings are unchanged (Table S1).

***Definitions of roaming and dwelling.*** In Fig. 3 we claim that the first two MDS dimensions correspond to a roaming and dwelling model. Here we describe how behaviors are classified into roaming and dwelling and show that these properties correlate with the MDS dimensions. For each individual we construct a full-lifetime histogram $\overline{P}^N = \langle P^N(t) \rangle$ where brackets denote an average over time. Two examples of such histograms are shown in Fig. S7 $A$ and $B$. We then fit a Gaussian mixture model of the form

$$\hat{\overline{P}}^N(|v|,\omega) = \sum_{i=1}^{5} \pi_i g_i\left(\mu_i^N, \Sigma_i^N\right), \qquad \textbf{[S6]}$$

where $g_i(\mu_i^N, \Sigma_i^N)$ is a bivariate normal distribution using expectation–maximization with unconstrained covariance. $\pi_i$ are the weights of $g_i$ such that $\sum_i \pi_i = 1$, $\mu_i^N = (|v|_0, \omega_0)_{N,i}$, and

$$\sum_i^N = \begin{bmatrix} \sigma_{|v|}^2 & \sigma_{|v|,\omega}^2 \\ \sigma_{|v|,\omega}^2 & \sigma_\omega^2 \end{bmatrix}_{N,i}, \qquad \textbf{[S7]}$$

where $N$ is an index over individuals. Information criteria (Bayesian or Akaike) indicate that five independent Gaussians are needed to provide a good fit to the data. We note that this fit could be improved by using a more appropriate mixture model (e.g., one that accounts for the fact that $|v|$ is strictly positive) but that it does provide a reasonable classification of our data. Also, we found that ad hoc thresholding procedures that detected minima in $\overline{P}^N$ give similar correlations between MDS dimensions and properties of the behavior.

We fit Eq. **S6** to $\overline{P}^N$, using expectation–maximization for all 171 individuals that were included in the individuality matrix embedded in Fig. 3. Fig. S7$C$ shows a scatter plot of $|v|_0$ and $\sigma_\omega$ for all $N$ and $i$. The $g_i$ with high $|v|$, low $\sigma_\omega$ correspond to roaming behavior as indicated in Fig. S7$D$. For each individual, we classify $g_i$ as roaming if $|v|_0 > 100\mu m/s$ and $\sigma_\omega < 7 rad/s$ (this results in two to three Gaussians for each individual). We designate the $g_i$ with the highest $|v|_0$ for each individual as the roaming state mode $g^{max}$.

For each behavior, $P^N(s)$, we then classify actions, i.e., $(|v|,\omega)$ pairs, as roaming or dwelling by clustering the data according to

the Gaussian mixture model ($\hat{\bar{P}}^N$). Examples of clustering results for two full-lifetime histograms are shown in Fig. S7 *A* and *B*. Using this classification, we calculate the fraction of the time each $P^N(s)$ spends roaming and this quantity is correlated with MDS dimension 1 in Fig. S7 *D* and *G*. Finally, for each $P^N(s)$, if more than 50% of the actions in a given behavior are clustered to $g^{max}$, we compute the mode of speed distribution in the roaming state. This amounts to finding the mode of the speed distribution for behaviors with sufficient density in the roaming state to specify this mode. The $|v|$ corresponding to the this mode is designated the "roaming speed" and is correlated with MDS dimension 2 in Fig. 4 *E* and *G*. The correlation for all behaviors from the embedding is shown in Fig. S7 *D* and *E* for WT *T. thermophila*:1xR and in Fig. S7 *F* and *G* for all conditions. Table S2 gives the correlation coefficients for the data in Fig. S7 *F* and *G*.

We find that the correlations between the first two MDS dimensions are qualitatively maintained irrespective of the type of MDS applied to the data.

***Hidden Markov models.*** As discussed in the main text, Gallagher et al. (15) pursue a hidden Markov model (HMM) strategy for describing behavioral recordings of *Caenorhabditis elegans* over periods of 1–4 h. By fitting many independent HMMs for different individuals in various environments and with various mutations, they find that the parameter space of HMMs is effectively low dimensional. We explored the possibility of describing behavioral variation in *T. thermophila* by a similar strategy. However, we found that the nonstationarity present in our behavioral measurements resulted in HMMs describing the behavioral variation poorly. In particular, by definition, the residence time distribution for each state in an HMM is exponential. We found that the HMMs inferred from our data produced state residence time distributions that differed qualitatively from the residence time distributions for *T. thermophila* (longer tails). There are methods for constructing HMMs with nonexponential residence time distributions; however, these models are ad hoc and require inferring a large number of parameters (16). As a result, we pursued the analytical method presented here. Moreover, inferring HMMs for different portions of an individual's lifetime resulted in qualitatively different transition matrices—supporting our claim that behavioral dynamics are nonstationary.

Ultimately, our results provide a route to a very similar behavioral characterization to that observed by Gallagher et al. (15); however, we do not require explicit specification of emission probability densities or inference of the number of hidden behavioral states. As a result our method is potentially more broadly applicable. It is striking, however, that we find an effectively similar set of behavioral states. It would be interesting to determine whether this three-state characterization of *C. elegans* behavior is maintained when full-lifetime behavioral variation is measured.


1. Sweet, MT, Allis CD (2006) Long-term storage of unfrozen Tetrahymena cultures in soybean medium. *CSH Protoc.*
2. Andersen AP, and Hellung-Larsen P (1989) Division competence in Tetrahymena: Determination of minimum cell volume and rate of nutrient uptake. *J Cell Biochem* 41(3):125–133.
3. Rasmussen L, Modeweg-Hansen L (1973) Cell multiplication in Tetrahymena cultures after addition of particulate material. *J Cell Sci* 12(1):275–286.
4. Szablewski L, et al. (1991) Tetrahymena thermophila: Growth in synthetic nutrient medium in the presence and absence of glucose. *J Eukaryot Microbiol* 38(1):62–65.
5. Frentz Z, Kuehn S, Hekstra D, Leibler S (2010) Microbial population dynamics by digital in-line holographic microscopy. *Rev Sci Instrum* 81(8):084301.
6. Quake SR, Scherer A (2000) From micro- to nanofabrication with soft materials. *Science* 290(5496):1536–1540.
7. Jaqaman K, et al. (2008) Robust single-particle tracking in live-cell time-lapse sequences. *Nat Methods* 5(8):695–702.
8. Jonker R, Volgenant A (1987) A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing* 38(4):325–340.
9. Winet H (1973) Wall drag on free-moving ciliated micro-organisms. *J Exp Biol* 59(3): 753–766.
10. Freedman D, Diaconis P (1981) On the histogram as a density estimator: L 2 theory. *Probab Theory Relat Fields* 57(4):453.
11. Grosse I, et al. (2002) Analysis of symbolic sequences using the Jensen-Shannon divergence. *Phys Rev E* 65(4 Pt 1):041905.
12. Endres DM, Schindelin JE (2003) A new metric for probability distributions. *IEEE Trans Inf Theory* 49(7):1858–1860.
13. Bourgain J (1985) On lipschitz embedding of finite metric spaces in Hilbert space. *Israel J Math* 52(1–2):46–52.
14. Matousek J (1996) On the distortion required for embedding finite metric spaces into normed spaces. *Israel J Math* 93:333–344.
15. Gallagher T, Bjorness T, Greene R, You Y-J, and Avery L (2013) The geometry of locomotive behavioral states in C. elegans. *PLoS ONE* 8(3):e59865.
16. Rabiner L (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE* 77(2):257–286.

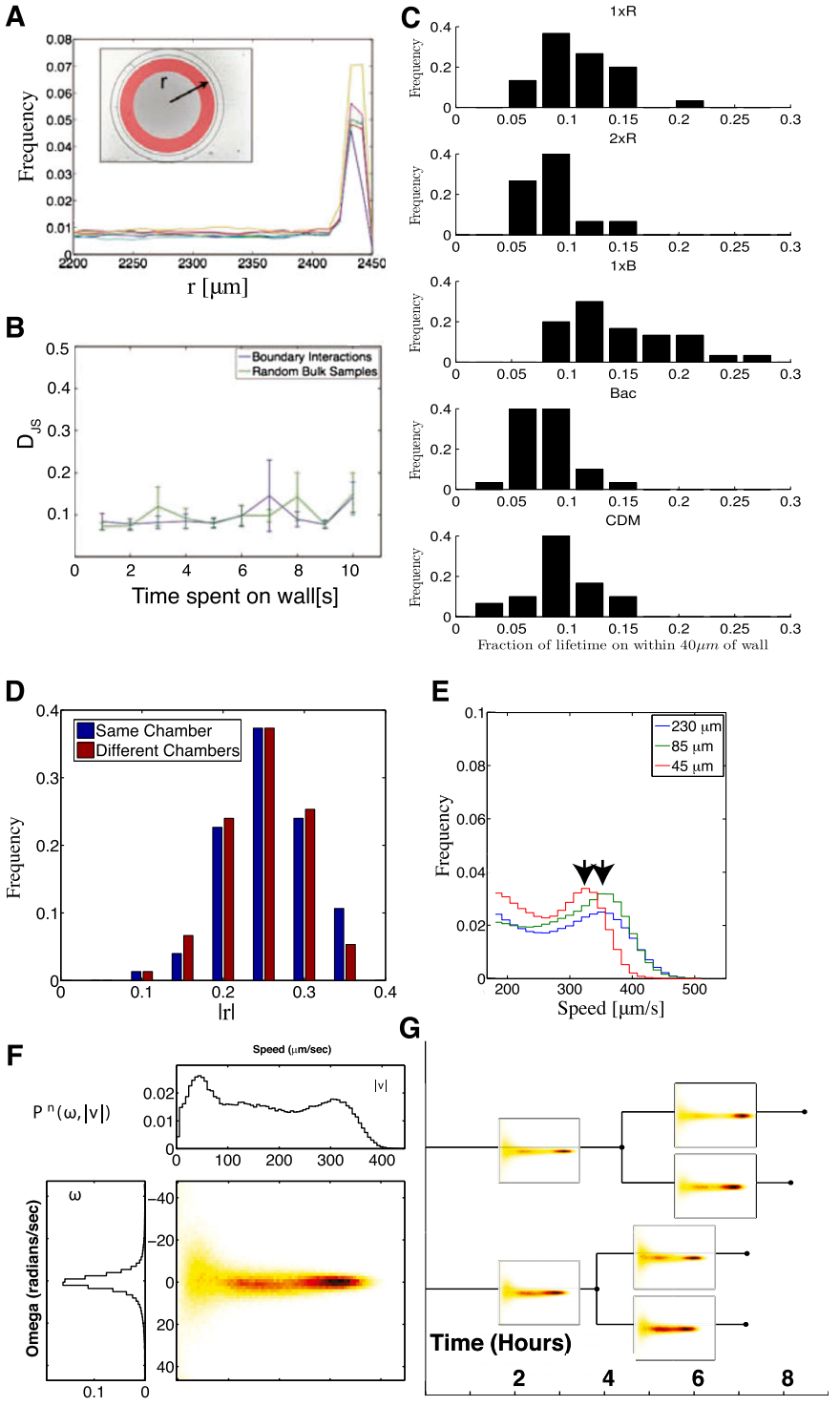



**Fig. S1.** Chamber homogeneity and wall interactions. (*A*–*C*) Wall interactions. (*A*) The radial average distribution of the centroid location as a function of the distance from the center of the chamber (*r*). The six traces are constructed from the full lifetimes of all individuals from a single experiment. Note the uniform distribution beyond ~40 μm from the wall, naturally defining a wall interaction region. The fraction of the total lifetime that an individual spends within this region varies between 5% and 30% across conditions as shown in *C*. The Bac individuals spend the largest fraction of their lifetime swimming rapidly and therefore the greatest time interacting with the walls. *B* shows that there is no lasting, measurable, effect on the speed and turning angle distributions of these wall interaction events. Jensen–Shannon divergence estimates are shown, using speed distributions alone, for short (10-frame) segments immediately before and after a wall interaction event of varying durations (blue). The same analysis was repeated for randomly chosen sections of trajectories where no wall interaction occurred (green). We conclude that there is no "memory" in terms of the speed distributions due to wall interaction events. (*D*–*F*) Chamber homogeneity. (*D*) We compute the 2D spatial distribution of centroid locations $(\rho^i_{xy})$ for individual *i* over its full lifetime. We then compute the correlation coefficient $r_{ij}=\frac{\sum_{xy}(\rho^i_{xy}-\rho^i_{xy})(\rho^j_{xy}-\rho^j_{xy})}{\sum_{xy}(\rho^i_{xy}-\rho^i_{xy})^2(\rho^j_{xy}-\rho^j_{xy})^2}$, where $\langle\cdot\rangle$ denotes a spatial average over *x* and *y*, for all pairs of individuals. We computed *r* for 36 pairs of individuals residing in the same chamber and an additional 36 pairs residing in different chambers. $\rho^i_{xy}$ was constructed by omitting the division portion of the lifetime,

Legend continued on following page

where cells swim slowly or not at all, and the wall interaction region. (*E*) Effect of "floor" and "ceiling" interactions. Histograms of swimming speed for three different chamber depths are averaged across $n = 18$ individuals (three families per chamber depth). Arrows indicate decreasing peak speed with decreasing chamber depth. (*F*) An example of a full-lifetime average histogram $\overline{(P^N)}$. (*G*) The phylogeny of an entire family (six WT *T. thermophila*:1xR individuals) including their full-lifetime average histograms. The black dots indicate division events and the lower axis, time.

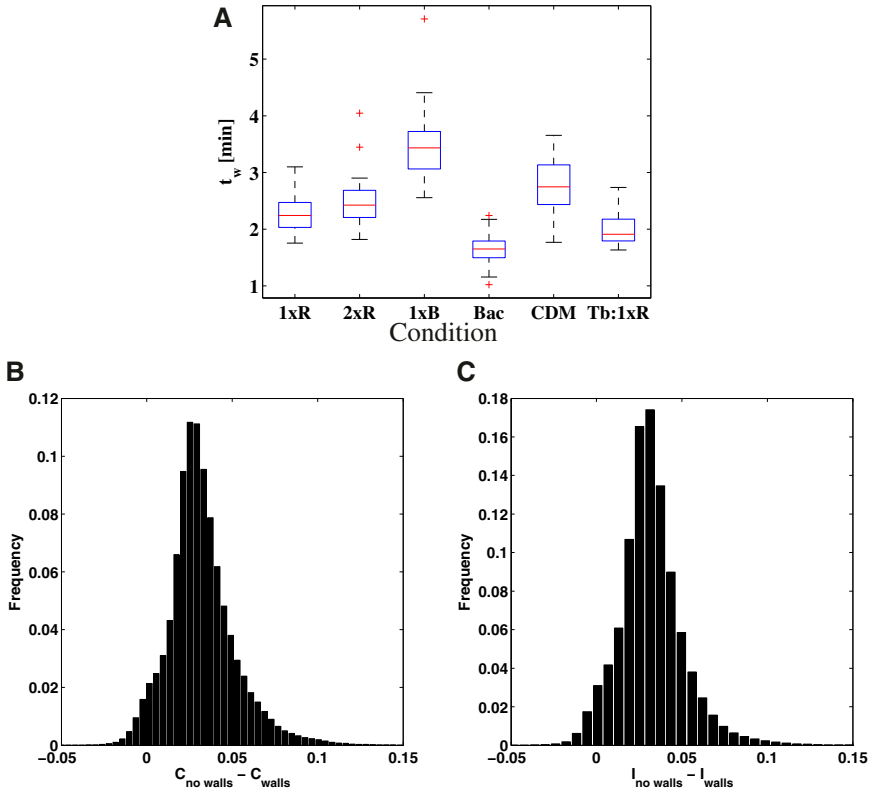



**Fig. S2.** Box plot of $t_w$ for $I^{N,M}(s,s')$ calculations and effects of wall removal. (*A*) Variable $t_w$. To compute the individuality matrices for Figs. 2 and 3 of the main text we must divide each individual's lifetime into $n = 100$ equal duration segments. As a result the length of these segments varies from one individual to the next. Here we show a box plot giving the distribution of window sizes, for all 30 individuals in each of five conditions. The comparatively long windows in 1xB are due to the long generation times in that condition. (*B* and *C*) Effect of wall removal on individuality and changeability. To measure the effect of wall interactions on our changeability and individuality matrices we recalculated these quantities with and without the wall interactions included (for changeability $C_{walls}$ and $C_{\text{no walls}}$, respectively) for $n = 50$. We then computed the difference between these matrices. (*B*) For all 171 individuals used in Figs. 3 and 4 the distribution of these differences is shown. We repeated the calculation for the individuality matrices between all 171 individuals, and the results are shown in *C*. Because these differences are small, we conclude that our estimates of changeability and individuality are not strongly influenced by cells interacting with the walls of the chamber.

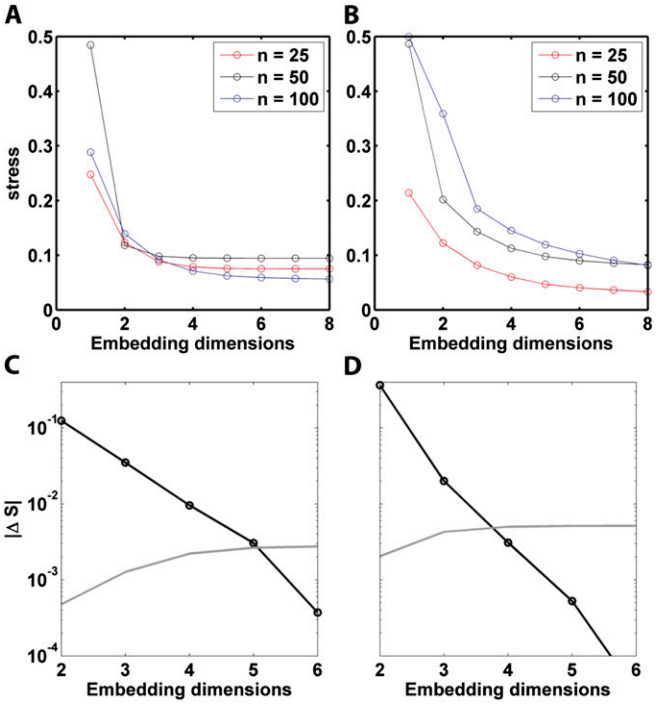



**Fig. S3.** Stress as a function of the number of embedding dimensions for $n = 25$, 50, 100 for all 171 individuals embedded in Fig. 3*H*. (*A*) Stress plots for $D_{JS}$. (*B*) Stress plots for $\sqrt{D_{JS}}$. For all embeddings our interpretation of the first two MDS dimensions is consistent. (*C* and *D*) The change in stress $|\Delta S|$ as a function of the embedding dimension for $n = 25$ (*C*) and $n = 50$ (*D*). All embeddings are performed on $D_{JS}$. Black lines with circles measure the decrease in stress by embedding in the number of dimensions indicated on the *x* axis. The gray line is Eq. **S5**, indicating the minimum significant change in stress. At smaller *n* the effect of noise is decreased because the number of points used to compute each $P^N(s)$ increases.

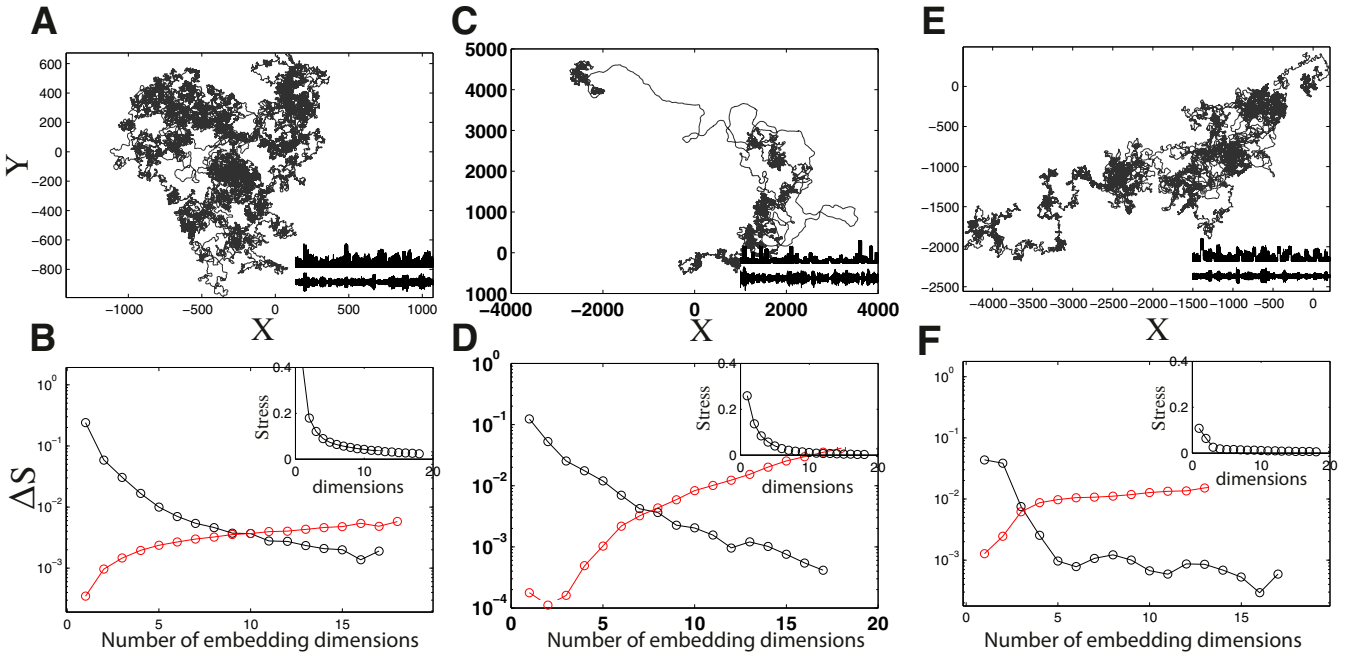



**Fig. S4.** Simulations demonstrate that low-dimensional embeddings are nontrivial. We simulated trajectories for three types of synthetic $P^N(|v|, \omega)$ histograms: a "high-dimensional" histogram with 11 unconstrained free parameters (*A* and *B*), the same histogram with a hyperbolic constraint applied (*C* and *D*), and a "low-dimensional" histogram with only 3 free parameters (*E* and *F*). *A*, *C*, and *E* show real-space trajectories, and *Insets* show $|v|$ and $\omega$ time series. *B*, *D*, and *F* show stress plots identical to those in Fig. S6 for the trajectories in *A*, *C*, and *E*, respectively. The red line is $|\Delta S_{min}|$ as in Fig. S6. Note that the low-dimensional trajectory embeds in fewer dimensions.

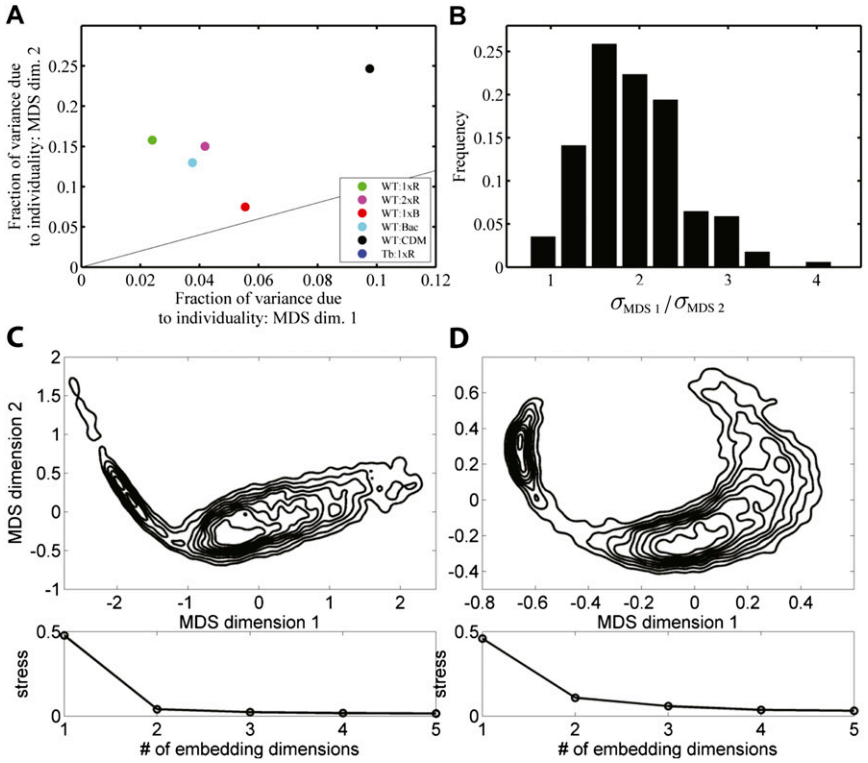



**Fig. S5.** For the embedding shown in Fig. 3*H*, changeability and individuality occur largely in MDS dimensions 1 and 2, respectively. (*A*) The fraction of variance in each of the two MDS dimensions due to individuality is calculated for the embedding in Fig. 3*H*. This fraction is plotted by condition for each of the five conditions where the WT *T. thermophila* strain was studied as well as for *T. borealis*. Because all points lie above the equality line, the variance due to individuality in MDS dimension 2 is larger than that in MDS dimension 1. (*B*) For each individual we compute a SD along each of the two MDS dimensions in the embedding shown in Fig. 3. A histogram of the ratio $(\sigma_{MDS1}/\sigma_{MDS2})$ of these SDs shows that for most individuals (98%) it is larger than 1, indicating that changeability occurs primarily in MDS dimension 1. *C* and *D* show MDS embeddings using other metrics to compare $P^N(s)$. (*B*) We computed a distance between $P^N(s)$ histograms, using the earth mover's distance (EMD), and then performed an embedding. Because computing the EMD is computationally expensive, we did this for 50 behaviors per lifetime for all 171 individuals. We then performed metric MDS on this dissimilarity matrix and the embedding in two dimensions is shown in *C* with the corresponding stress plot (*Lower*). Again, we find that two dimensions capture the variation across the population well. As a second check, we computed the distance between histograms of speed only (1D histograms), using the maximum distance between the cumulative distribution functions [Kolmogorov–Smirnov (KS) test statistic]. (*D*) Using this metric to measure distances between behaviors, we computed a dissimilarity matrix and embedded the matrix using metric MDS. Again we found that two dimensions were sufficient to represent the diversity represented by 50 behaviors per lifetime for all 171 individuals. For both the EMD and KS measurements of histogram dissimilarity the correlations between the two MDS dimensions and our two-state model of behavior hold.

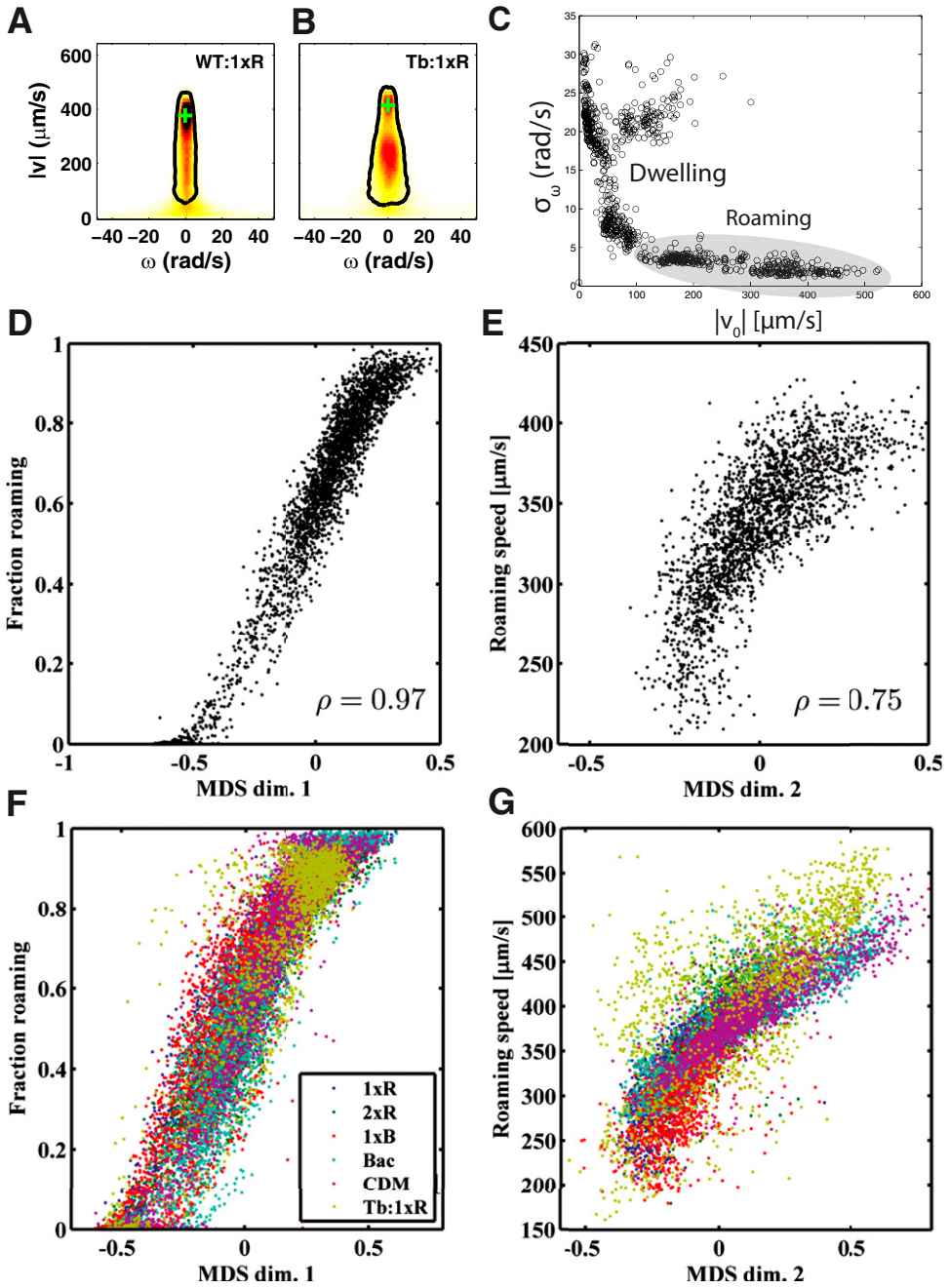



**Fig. S6.** Correlation between MDS dimensions and two-state model. (*A* and *B*) Full-lifetime average histograms $\overline{(P^N)}$ for two individuals [WT *T. thermophila*:1xR and *T. borealis* (Tb):1xR, respectively]. The density within the black line demarcates the roaming state and the green "+" sign shows the location of the roaming speed. (*C*) A scatter plot of $v_0$ and $\sigma_\omega$ for Gaussian mixture model fits to all 171 individuals considered in Fig. 3 (all *i* and *N*, see *Definitions of roaming and dwelling* section for definitions). The roaming and dwelling regions are labeled. (*D*) A scatter plot of the high-state residence time and the projection along MDS dimension 1 for 30 WT:1xR individuals in Fig. 3*A*. The correlation coefficient $(\rho)$ is shown. (*E*) A scatter plot of the location of the high-state mode and the projection along MDS dimension 2 for the same population. (*F*) Same as *D* except for all six populations from the full embedding in Fig. 3*H* of the main text; applies to *G* as well. (*G*) Same as *E* except for all six populations of WT *T. thermophila* and *T. borealis*. Correlation coefficients are listed in Table S2.

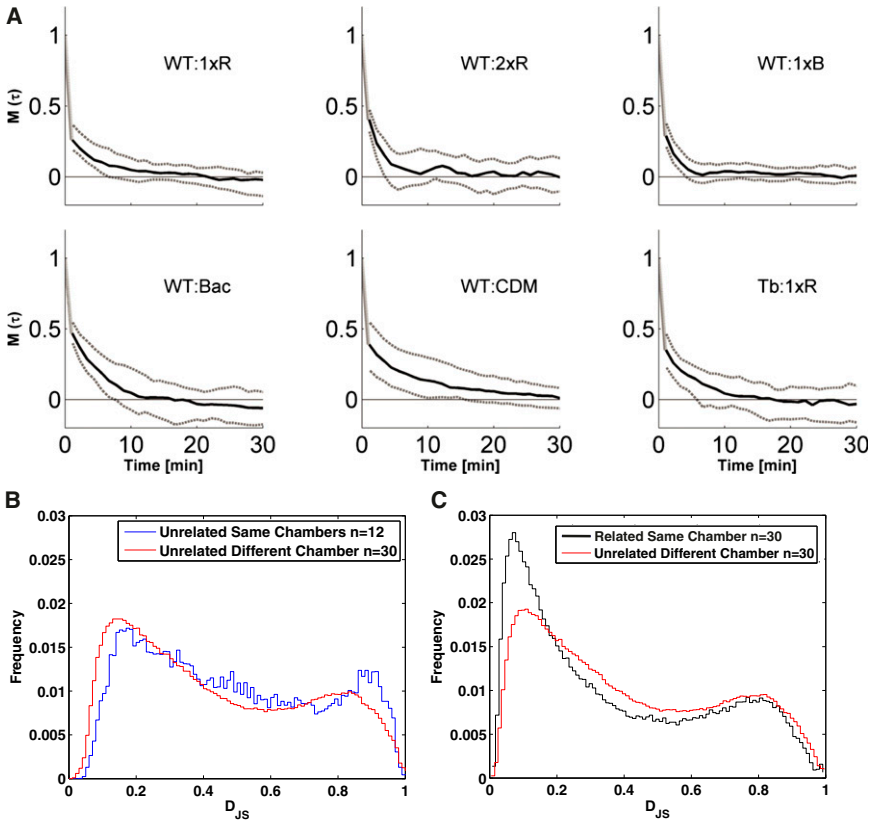



**Fig. S7.** (*A*) Memory plots identical to those shown in Fig. 4 *E* and *F* for all six populations studied. Each plot is labeled by strain:condition. For all five WT *T. thermophila* strains each plot is composed of 30 individuals, and that for *T. borealis* (Tb) is composed of 21 individuals. The black line indicates the median of the population and the gray lines the 0.1 and 0.9 quantiles. (*B* and *C*) Heritability control experiment. (*B*) To exclude the possibility that the heritability effects we observe are due to environmentally specific heterogeneity we placed two unrelated WT *T. thermophila*:CDM individuals in the same chamber and measured their behavior. The distribution of $D_{JS}$ distances for unrelated WT *T. thermophila*:CDM individuals in the same chamber is compared with the same measurement for unrelated individuals in different chambers. Because unrelated individuals in the same chamber are slightly less similar than unrelated individuals in different chambers, we conclude that the heritability we observe does not arise from related individuals residing in the same chamber. (*C*) To demonstrate that distributions of heritable behaviors are overlapping we show distributions of all $D_{JS}$ values for related and unrelated WT *T. thermophila*:CDM individuals.

**Table S1. Analysis of MDS embeddings for different types of scaling and various *n*, with and without $\sqrt{D_{JS}}$**

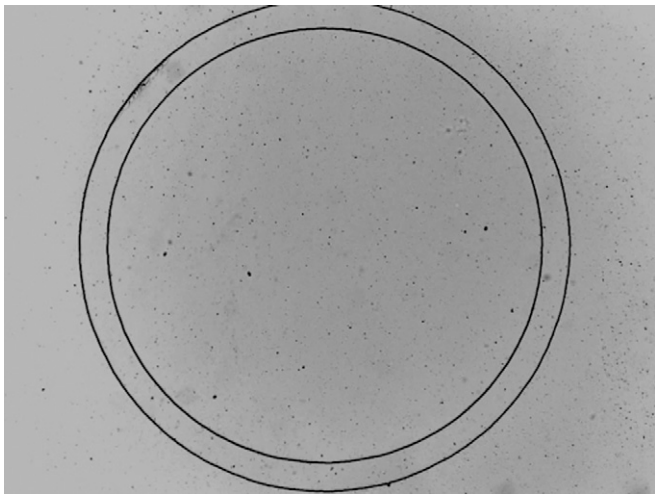
| *n* | $\sqrt{D_{JS}}$? | MDS type | $\sigma^2_{MDS2}$ Ind. | $\sigma^2_{MDS1}$ Ind. | $\sigma_{MDS1}/\sigma_{MDS2} > 1$ | Stress |
|---|---|---|---|---|---|---|
| 25 | No | Nonmetric | 0.42 | 0.1 | 0.96 | 0.12 |
| 25 | Yes | Metric | 0.42 | 0.11 | 0.84 | 0.22 |
| 25 | No | Metric | 0.42 | 0.1 | 0.96 | 0.12 |
| 50 | No | Nonmetric | 0.35 | 0.1 | 0.99 | 0.11 |
| 50 | Yes | Metric | 0.36 | 0.1 | 0.83 | 0.20 |
| 50 | No | Metric | 0.36 | 0.1 | 0.98 | 0.11 |
| 100 | No | Nonmetric | 0.31 | 0.1 | 0.91 | 0.12 |
| 100 | Yes | Metric | 0.31 | 0.1 | 0.8 | 0.22 |
| 100 | No | Metric | 0.31 | 0.1 | 0.95 | 0.12 |
| 50 | EMD | Metric | 0.38 | 0.13 | 1 | 0.04 |
| 50 | KS | Metric | 0.28 | 0.07 | 0.8 | 0.1 |

$\sigma^2_{MDS2}$ Ind. is the fraction of variance in MDS dimension 2 due to individuality and likewise for $\sigma^2_{MDS1}$. Because a larger fraction of variance in MDS dimension 2 is due to individuality, this dimension is consistently associated with individuality. $\sigma_{MDS1}/\sigma_{MDS2} > 1$ is the fraction of individuals for which the variance in MDS dimension 1 exceeds that in dimension 2. Because this fraction is always large, changeability resides largely in MDS dimension 1. The bottom two rows show results for other metrics: the earth mover's distance (EMD) and the Kolmogorov–Smirnov test statistic (KS) on cumulative distribution functions of $|v|$ alone.

**Table S2. Behavioral memory, mean lifetimes, and correlation between changeability and individuality and MDS dimensions**

| Condition | Memory, min | Average lifetime, min | Memory/lifetime | $\rho 1$ | $\rho 2$ |
|---|---|---|---|---|---|
| 1xR | 22 | 190 | 0.12 | 0.96 | 0.82 |
| 2xR | 30 | 106 | 0.28 | 0.97 | 0.78 |
| 1xB | 27 | 294 | 0.09 | 0.91 | 0.61 |
| Bac | 16 | 129 | 0.12 | 0.95 | 0.86 |
| CDM | 34 | 197 | 0.17 | 0.95 | 0.86 |
| Tb:1xR | 17 | 115 | 0.15 | 0.92 | 0.73 |

Memory is computed as the first zero crossing of the population's median memory. The last two columns show correlations between MDS dimensions and two-state model correlations. The correlations shown correspond to the embedding in Fig. 3*H*. $\rho_1$ is the correlation coefficient between MDS dimension 1 and the fraction of time spent roaming, and $\rho_2$ is that between MDS dimension 2 and the roaming speed. Correlations depend quantitatively, but not qualitatively, on the type of MDS used or whether we embed $\sqrt{D_{JS}}$ or $D_{JS}$. Tb, *T. borealis*.



**Movie S1.** An unprocessed movie of four WT *T. thermophila* (TtWT):1xR individuals in a microfluidic chamber like the one shown in Fig. 1*A*.

Movie S1

**Movie S2.** A background-subtracted version of Movie S1.

Movie S2